

### Extraction (out) of subjects, objects and predicate nominals

Long-distance extractions show strong statistical asymmetries between subjects and objects, modulated in important ways by construction type. Wh-questions prefer object extraction over subject extraction, both in Dutch and in English, whereas relativization and topicalization prefer subject extraction (Schippers & Hoeksema 2021). The latter suggest that topicality versus focus may explain these differences. Wh-expressions are focus elements, whereas topicalized expressions and relative pronouns have topic status (cf. Bayer 2005; Mak et al. 2008). Topicality is associated with subjects, not with direct objects, making subjects most suitable for relativization and topicalization, but less suitable for wh-movement.

For extraction out of subjects and objects, often a similar asymmetry is proposed: extraction out of direct objects is fine, extraction out of subjects is ruled out (cf. Huang 1982). Such distinctions are based on limited data, such as:

- (1) a. \*Who did a brother of kiss Mary?  
b. Who did Mary kiss a brother of?

This contrast is actually not strong: both sentences are degraded. Also, many studies ignore any differences between distinctive types of long-distance extractions.

We present corpus data from the Dutch Lassy Large corpus involving the words *waarvan* and *waar* + postposition, which can be used as wh-items and relative pronouns. *Waarvan* is much more likely to be a relativizer than a question word, and much like *dont* in French, may easily be associated with subjects as an attributive modifier (Tellier 1990, Abeillé & Winckel 2020). The data for *waar* show an almost complete absence of attributive uses, and many cases where *waar+postposition* forms another kind of unit, often a prepositional object or a PP adverbial. The absence of attributive uses is akin to cases like (1) in English, but are not limited to subjects. This would fit in with a ban of R-word extraction out of PP-in-NP, but not with a subject condition. We compare our results with data from English and French.